

biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

Visitor-Art Interaction by Motion Path Detection

Viktor Slavkovikj, Pieterjan De Potter, Steven Verstockt, Wesley De Neve, Sofie Van Hoecke, and Rik Van de Walle

In: IEEE International Conference on Digital Signal Processing (DSP 2013), 2013.

To refer to or to cite this work, please use the citation to the published version:

Slavkovikj, V., De Potter, P., Verstockt, S., De Neve, W., Van Hoecke, S., and Van de Walle, R. (2013). Visitor-Art Interaction by Motion Path Detection. *IEEE International Conference on Digital Signal Processing (DSP 2013)*

Visitor-Art Interaction by Motion Path Detection

Viktor Slavkovikj^{*†}, Pieterjan De Potter^{*}, Steven Verstockt^{*},
Wesley De Neve^{*‡}, Sofie Van Hoecke^{*†}, and Rik Van de Walle^{*}

^{*} Multimedia Lab, Department of Electronics and Information Systems,
Ghent University-iMinds, Ghent, Belgium

[†] Electronics and Information Technology (ELIT) lab,
Ghent University campus Courtray, Courtray, Belgium

[‡] Image and Video Systems Lab, Korea Advanced Institute of Science and Technology (KAIST),
Yuseong-gu, Daejeon, 305-732, Republic of Korea

{viktor.slavkovikj, pieterjan.depotter, steven.verstockt, wesley.deneve, sofie.vanhoecke, rik.vandewalle}@ugent.be

Abstract—This paper describes a method for video-based motion path detection which is applied in the creation of an interactive artwork. The proposed algorithm, based on the Hough transform, detects parametric motion trajectories in real-time (10 fps). In order to detect people's motion under non-static background object occlusion we have also developed a video segmentation technique. The proposed interaction system adopts top-down camera view to extract spatiotemporal motion trajectories and discern predefined patterns of movement thus enabling the creation of new artistic choreographies. We present test results that illustrate the effectiveness of our method and discuss the practical applicability of our approach in other domains.

Index Terms—motion path detection; real-time video processing; video segmentation

I. INTRODUCTION

Contemporary art is often influenced and augmented by technology. In this sense, the expressive power offered by different multimedia technologies is particularly useful in the creation of interactive art. An overview of interactive artistry, employing a variety of multimedia methods, is given by Jaimes et al. [1], [2], and Tinapple and Ingalls [3].

This paper deals with the analysis of the motion of people from video which has been applied in the realization of an interaction artwork. Figure 1 shows a diagram of the artistic setup. The conceptual idea of the artist in this kinetic installation is to translate urban data, weather conditions, and the motion of visitors into a dramatic choreography. To observe and influence the creation of emerging and disappearing visual and acoustic patterns, visitors can enter under a synthetic cloud formed by the swirling and circling movement of a large transparent sail. We analyze the video sequences obtained from static cameras placed in top-down view. This enables us a view of the scene (without having visitors occluding each other) from which we can extract motion trajectories. The extracted motion trajectories are used to infer parametric patterns of

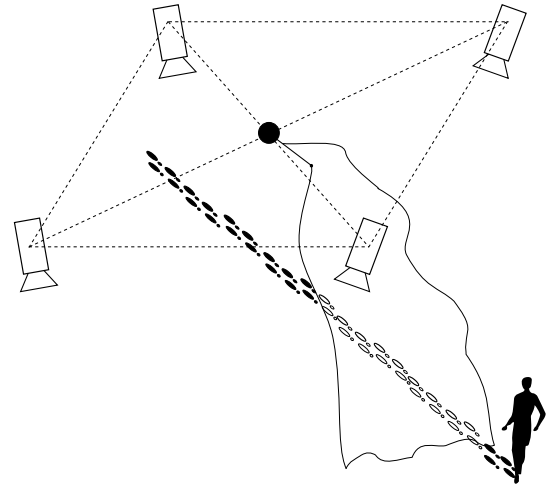


Fig. 1: A diagram of the interactive art exhibition system.

observers' movements. Detected motion patterns are utilized in the creation of new sail choreographies in real-time.

The remainder of this paper is organized as follows. Section II is devoted to related work in motion trajectory analysis and shape detection. In Section III, we propose a novel algorithm for observers' motion path detection. Next, Section IV gives an overview of the architecture of the artistic system. The obtained results are described in Section V. Finally, we conclude this paper in Section VI and discuss other possible applications of the proposed method.

II. RELATED WORK

Motion trajectories are features of interest in many automatic video surveillance applications. Detection of motion paths is of particular importance in anomalous behavior identification, where the tracked objects usually follow a recurrent movement path.

Makris and Ellis [4] model frequently used pedestrian paths from video footage of outdoor scenes for the purpose of long period logging of movement patterns, recognition of atypical movement, and probabilistic prediction of the route chosen by a pedestrian entering the scene. Here, trajectories are represented as resampled 2D points and routes of typical movement are detected by merging trajectories that satisfy a predefined minimal distance.

Sillito and Fisher [5] represent the problem of identifying anomalous behavior in pedestrian motion trajectories as a one-class learning problem. Each spatiotemporal motion trajectory is encoded as a vector of the control points (obtained by least-squares fitting) of a uniform cubic B-spline curve approximation of the original trajectory. A one-class classifier [6], based on Gaussian mixture model is trained on normal examples of motion trajectories. When a low likelihood example is encountered, approval of a human operator is required in order to incorporate it into the existing model or to label it as anomalous motion.

Typical pedestrian motion patterns are described by a probabilistic model in the work of Ellis et al. [7]. They build a Gaussian Process model for each group of training trajectories, clustered by start point, by inferring the model parameters from the observed data. The built model can then be used to give a future estimate of a target's position.

In the method of Ricci et al. [8] for learning common pedestrian trajectories, motion paths are represented by flow vectors, which consist of the position and velocity of the targets at a given time. A grouping of the aforementioned representation of trajectories is then performed by using a Kernel K-means with a Dynamic Time Warping kernel [9]. Kernel Canonical Correlation Analysis [10] is used on a training data set to learn a mapping function between positions of clustered trajectories and corresponding instantaneous velocity vectors. Finally, the learned mapping function is utilized to predict the velocity vector for a novel input trajectory position.

Apart from motion trajectory analysis, another important aspect related to our work is detection and recognition of shapes from images. The detection of shapes from image data is one of the basic problems in computer vision, as it relates to many image and video analysis applications.

In the work of Su et al. [11] detection of predetermined shapes in 2D point clouds, extracted from images, is formulated as a generalized likelihood ratio test. The points from the point cloud belonging to the predefined shape are assumed to be noisy versions of the realization of a one-dimensional Poisson process, while for the clutter points a two-dimensional Poisson process model is used. The ratio test is formed from the likelihood of a given point cloud being only clutter to the likelihood of the same point cloud being associated with a given shape class. The parameters for both of the likelihood functions are obtained using maximum likelihood estimation.

A shape detection method for images can be found in the work of Garlipp and Müller [12]. Their approach is based on rotational difference kernel estimators for detecting linear edges and circles.

III. MOTION PATH DETECTION

For the purpose of creating an interactive art experience, we are interested in the detection of visitors' motion patterns. In particular we would like to be able to detect when a visitor's motion trajectory forms a parametrized shape, i.e., a line or a circle in the scene. There are two main differences of our work compared to the approaches described in Section II. First, in typical motion path analysis applications, the trajectories taken by people can be of arbitrary form, however, the form of the frequent, or usual, trajectories is constrained by the scene infrastructure (such as sidewalks, crossings, passages etc). Here we are interested in detecting patterns of movement, such as lines and circles, that are frequently occurring in the motion of the visitors in an unconstrained scene instead of arbitrary motion trajectories. Second, unlike the methods discussed in Section II, shape detection in motion trajectory data should meet real-time video processing requirements for speed and be robust to missing information.

A block diagram of our approach is given in Figure 2. First we perform detection and segmentation on the frames of the video sequence in order to segment the sail in the video of the scene. Then we extract motion trajectories. Finally, the extracted paths are used to determine if some visitor has traversed the scene with a regular motion trajectory. In the next subsections we describe the components of the system in more detail.

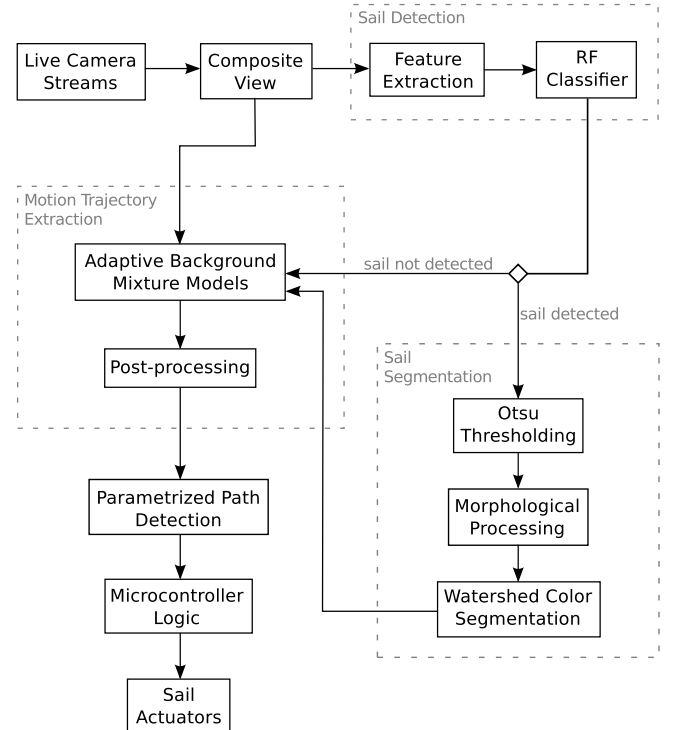


Fig. 2: Block diagram of the proposed system.

A. Sail Segmentation and Motion Trajectory Extraction

The presence of the sail in the field of view poses a challenge for the trajectory extraction of visitors' motion

patterns because the movement trajectory of the sail could be erroneously detected by the system together with the valid trajectories of the people moving in the scene. Also, due to the nature of the application (dimensions of the sail), it is not possible to position the cameras in such a way that their view would be unobstructed by the sail. Therefore, it is necessary to segment the sail in the video so that it can be included in the scene background. Applying segmentation of the sail to every frame of the video would not yield good results. This is because in the cases where the sail is out of the cameras' view, valid foreground objects could be segmented instead. We, therefore, approach the problem by first detecting whether the sail is present in the scene, and then performing segmentation only for those frames of the video for which the detection method gives a positive output.

1) *Sail Detection*: In order to detect whether the sail is present in a frame, we construct a binary classifier based on the random forest algorithm [13]. Random forest is an ensemble of individual decision tree predictors. Each decision tree is trained with the same parameters but on a random subset of the training set. These subsets are generated from the training set by using a bootstrap procedure. That is, for each tree in the collection you randomly select the same number of feature vectors as in the original training set. The selection of feature vectors is done with replacement i.e. for a given subset some vectors will occur more than once in the subset and some will not appear at all. In general, when the training set for a decision tree is constructed by drawing samples with replacement from the original training set, about two thirds of the observations will be included and one third will be left out. The observations that are left out are called out-of-bag data and are used for the internal estimation of the training error. When building the individual decision trees based on different random subsets of the original training data, at each node of each tree a random subset of the available variables is used to select how to best partition the dataset at that node. A new subset of variables is generated for every node, however, the size of the subset is fixed for all the nodes and trees. No pruning is performed i.e., each of the decision trees is built to its maximum size. The trained decision trees represent the final ensemble. The classification is performed so that first the input feature vector is classified by each of the individual tree models separately, then the output class label is obtained from the majority of all class labels. The randomness introduced in the selection of the training subsets and in the variable selection gives considerable robustness of the random forest model to outliers, overfitting, and noise [13].

To be able to design suitable features for the random forest classifier we have analyzed the graylevel histogram distribution of the image. We apply a Savitzky-Golay [14] smoothing filter to the image histogram to obtain filtered histogram data which is used as a feature vector in the random forest classifier. The filtering of the histogram data is order preserving, which allows to distinguish the prominent peaks in the histogram corresponding to uniform objects or areas in the image. The case where the sail is present in a frame

approximately results in a bimodal histogram distribution of the given frame.

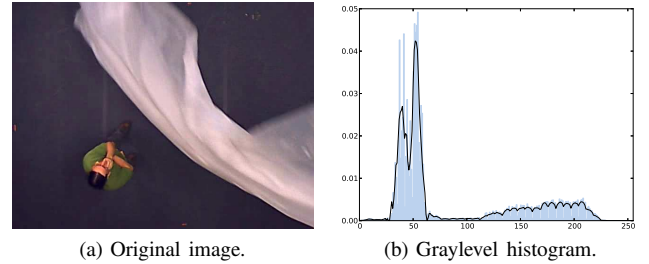


Fig. 3: Video frame containing the sail (a) and its graylevel histogram (b). The output of the Savitzky-Golay filter is shown on top of the histogram plot in (b).

In order to train the sail classifier, we built a dataset by uniformly sampling frames from a video sequence of a typical theater scene where the system was put to work. The video contained frames where the sail was included in the cameras' view and where it was not observed, both with and without people walking in the scene. Then we divided the acquired dataset into a training set containing approximately 2800 observations and a validation and test sets containing around 700 examples each. We trained the random forest classifier on the training set and used the validation set to obtain the optimal parameters by doing a grid search over parameter space. We tested the performance of the trained classifier on the separate test set for which we obtained a prediction accuracy of 95.55%.

2) *Segmentation*: A rough segmentation of the sail can be obtained by binarization of the video frames according to a predefined threshold. We use Otsu's method [15] to calculate this threshold automatically. This initial segmentation does not give sufficient results, however, it can be used as a basis to obtain an improved segmentation. The watershed transform [16] is a region-based segmentation method for gray-scale images which considers the gray-scale intensity as altitude in the topographic surface defined by the image. Areas in the image form catchment basins divided by watershed lines which are formed on the border of neighboring basins. Flooding the catchment basins starting from the minima segments the image in regions. The watershed algorithm depends strongly on the choice of local minima i.e., selecting many flooding sources can lead to over-segmentation of the resultant image. This is usually the case when using the gradient image to calculate the watershed transform. By introducing a color pixel similarity measure, the watershed algorithm can be extended for the segmentation of color images [17]. The flooding of the topographic surface of the image can also be treated as a region growing process from preselected seed pixels. We use the color image watershed segmentation with region growing seeds obtained from the morphological processing of the initial sail segmentation. The results of the segmentation using this approach can be seen in Figure 4.

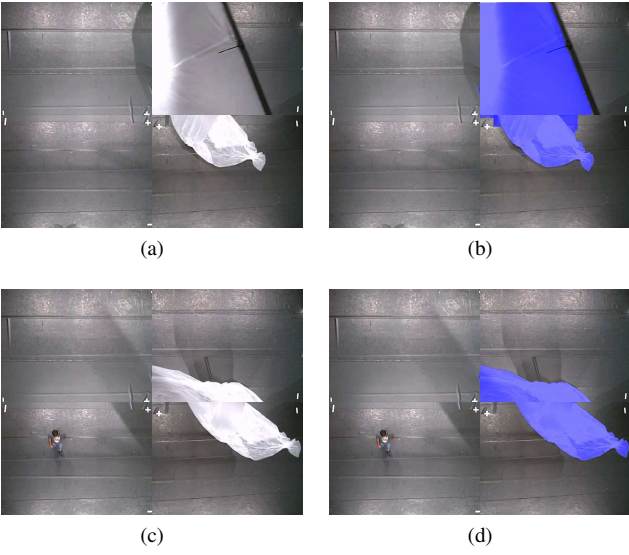


Fig. 4: Segmentation of the sail from frames in the video (left column). The segmented region is marked with false color (right column).

3) *Motion Trajectory Extraction*: In order to segment the moving targets from the background in the scene, we use adaptive background mixture models [18]. In this approach, each of the pixel values over a period of time are modeled as a mixture of Gaussians. A new pixel value is matched against the existing distributions. If a match is found, the parameters of the

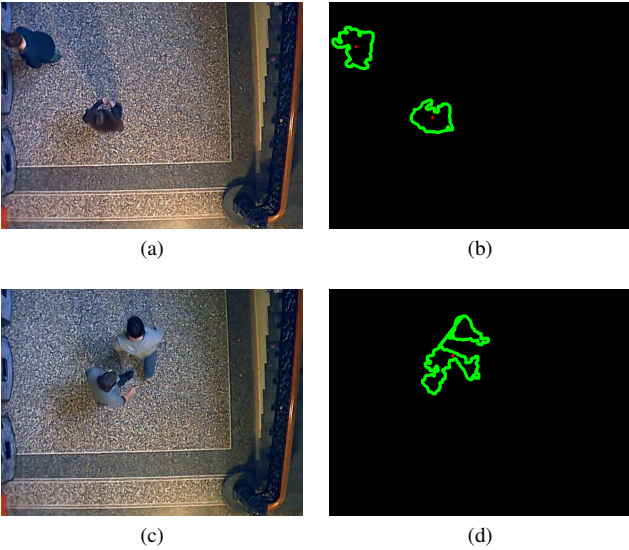


Fig. 5: Tracking targets on a complex background. Extracted contours with centroids (b) and (d) of targets in corresponding scenes (a) and (c). Two targets moving together are merged after detection (c) and (d).

distribution which matches the new observed pixel value are updated. Conversely, if a matching Gaussian is not found, then the least probable distribution is replaced with a new Gaussian distribution with mean equal to the current pixel value, having

low prior weight and a high variance. Gaussian distributions from the mixture that have high variance and high support from the observed data are considered to represent background processes. We incorporate the sail segmentation algorithm into the adaptive background mixture models method so that the pixels belonging to the moving sail are included in the background model.

We also employ post-processing techniques, namely filtering the segmented image sequences with a median filter, as well as applying morphological opening, to get rid of the noise from the previous background subtraction step. Figure 5 shows an example of the extracted contours of people moving in a area with a complex background, as detected by the algorithm. In order to approximate the position of the moving targets in the scene, we calculate the center of mass for each detected target. Note that detected objects that are moving together in a close group are merged, and their position is represented by a single centroid.

B. Parametrized Path Detection

The output from the previous step of our method gives the spatiotemporal position of each detected target in the scene (see Figure 6). Due to the fact that there may be

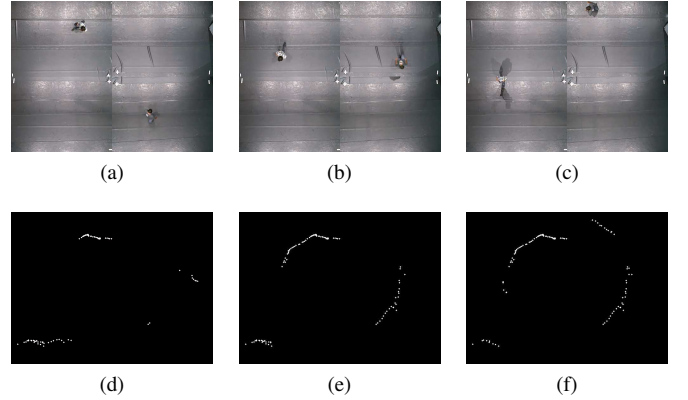


Fig. 6: Spatiotemporal motion trajectories (d)-(f) from the targets detected in the video of a scene (a)-(c).

inconsistencies in the tracking, such as when the tracked object is lost in a frame, or when the occlusion of the sail obstructs the detection of the target, it is necessary to account for noisy or missing data during the motion path detection. The standard Hough transform [19] is a robust statistical method for shape analysis in image data, which has been investigated and extensively utilized in image processing and computer vision before e.g., in the work of Illingworth and Kittler [20], and Kälviäinen et al. [21].

For line detection in images, the Hough transform maps a point (x, y) from the image plane to a set of points $\{(\rho, \theta) \mid \theta \in [0, \pi]\}$ in the Hough plane, according to the relation:

$$\rho = x \cos \theta + y \sin \theta. \quad (1)$$

Sinusoidal curves in the Hough plane, which represent different collinear points from the image plane, intersect at a point

(ρ', θ') giving the parameters of the line in the image plane in polar form. We augment the Hough line detection method so that line segments are detected in the spatiotemporal motion trajectory space. Short line segments are filtered out to prevent spurious path detection.

The Hough transform method can also be used for the detection of other parametrized shapes. We use a similar modification of the Hough transform algorithm to be able to find circular motion trajectory patterns. The relations that govern the transformation to Hough parameter space are given by:

$$a = x - r \cos \theta \quad (2)$$

$$b = y - r \sin \theta, \quad (3)$$

where a , b , and r denote the circle center position and radius, respectively, in parameter space. Each of the points lying on a circular pattern in the motion trajectory space forms a circle in Hough transform space. The circles intersect in the point (a, b) , which represents the center of the detected circle.

When detecting motion patterns, we utilize only the spatial positions for targets within a recent history time window in order to be able to detect the relevant patterns in real-time. Additional parameters, such as orientation and direction of movement, are calculated for each motion pattern together with the number and average speed of movement of the targets detected.

IV. OVERALL SYSTEM ARCHITECTURE OF THE ARTWORK

We constructed a client-server architecture for the interaction system. The video feeds of the four cameras in the camera network are processed with the previously described methods at the server. The parameters of the detected motion patterns are then sent to a microcontroller device. We use the Open Sound Control (OSC) [22] protocol for passing the data streams between the server and the microcontroller over the local area network. Based on the received parameters, the microcontroller logic creates new sail choreographies and sends control signals to the sail actuators.

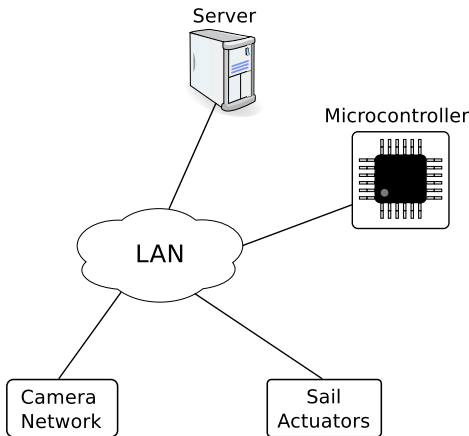


Fig. 7: Diagram of the client-server architecture of the system.

V. RESULTS

The results described in this section were obtained from a video sequence of a theater scene captured with the system. We used four consumer grade RGB color cameras for the camera array, which we placed in a top-down view to detect the motion patterns of the people moving below. The cameras were visually aligned during setup to provide a composite view of the scene. Note that additional techniques for registration of views are not necessary since small misalignments between camera views are accounted for in the design of the pattern detection algorithms.

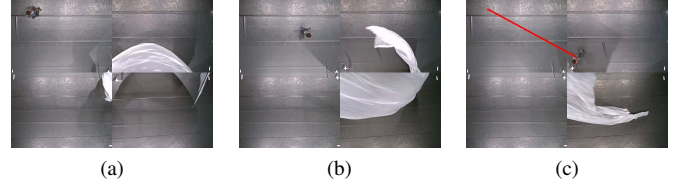


Fig. 8: Example of a person moving across the scene. The motion pattern detected by the system is shown in red.

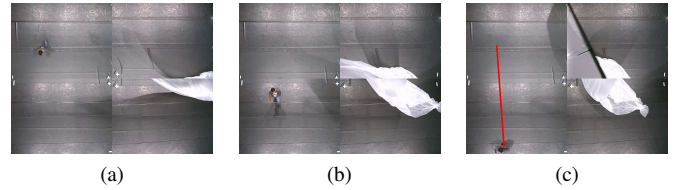


Fig. 9: A similar example to the one shown in Figure 7 with the target moving in a different direction.

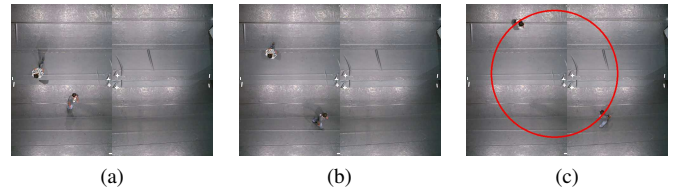


Fig. 10: Two targets moving in a circular pattern. The motion pattern detected by the system is shown in red.

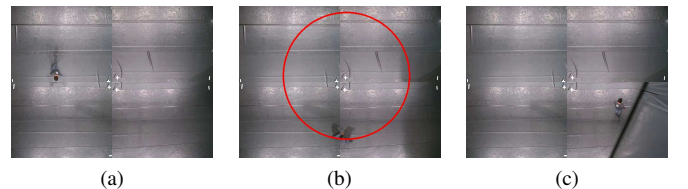


Fig. 11: Example of a circular motion pattern.

We compared the results of the system with the ground truth results of the test video sequence. The total duration of the test video was 11 minutes. During this time two participants were walking in the scene. Out of the 55 patterns made by the

participants, 42 were correctly detected by the system, there were 13 false negatives and 4 false positives, which gives a precision of 91.3% and a recall of 76.4% for the system.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a method for motion path detection from video. The proposed method is used in the realization of an interactive artwork, where visitors' motions are translated into novel artistic choreographies. The contributions of our work are twofold. We have developed an automatic algorithm which can infer predefined patterns of movement in real-time. We have also designed a video segmentation technique that allows the described system to operate under occlusions.

Detection of additional parametrized motion trajectories can be easily incorporated into the existing system. Other applications of the described technique are also possible. One application is the analysis of the rigid body motion of performance dancers without the aid of markers. Another relevant domain is traffic monitoring and analysis, which stands to gain a great deal from machine vision. To this end, our method could be used in pedestrian alerting systems, such as the one described in the work of Zhao et al. [23], and for vehicle counting and assessment of road utilization. Finally, the proposed approach could also be applicable in crowd analysis e.g. to avoid crowd related disasters [24].

ACKNOWLEDGMENTS

The research activities as described in this paper were funded by Ghent University and the Independent Research Institute iMinds.

REFERENCES

- [1] A. Jaimes and P. Jennings, "ACM multimedia interactive art program: an introduction to the digital boundaries exhibition," in *Proceedings of the 12th annual ACM international conference on Multimedia*, ser. MULTIMEDIA '04. New York, NY, USA: ACM, 2004, pp. 979–980.
- [2] A. Jaimes, A. Senior, and W. Muench, "ACM multimedia interactive art program: an introduction to the presence/absence exhibition," in *Proceedings of the 13th annual ACM international conference on Multimedia*, ser. MULTIMEDIA '05. New York, NY, USA: ACM, 2005, pp. 1051–1052.
- [3] D. Tinapple and T. Ingalls, "ACM multimedia interactive art program: interaction stations," in *Proceedings of the 19th ACM international conference on Multimedia*, ser. MM '11. New York, NY, USA: ACM, 2011, pp. 593–594.
- [4] D. Makris and T. Ellis, "Path detection in video surveillance," *Image and Vision Computing*, vol. 20, no. 12, pp. 895 – 903, 2002.
- [5] R. Sillito and R. Fisher, "Semi-supervised learning for anomalous trajectory detection," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2008, pp. 103.1–103.10.
- [6] R. R. Sillito and R. B. Fisher, "Incremental one-class learning with bounded computational complexity," in *Proceedings of the 17th international conference on Artificial neural networks*, ser. ICANN'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 58–67.
- [7] D. Ellis, E. Sommerlade, and I. Reid, "Modelling Pedestrian Trajectories with Gaussian Processes," in *Ninth International Workshop on Visual Surveillance*, 2009.
- [8] E. Ricci, F. Tobia, and G. Zen, "Learning Pedestrian Trajectories with Kernels," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ser. ICPR '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 149–152.
- [9] M. Cuturi, J.-P. Vert, O. Birkenes, and T. Matsui, "A kernel for time series based on global alignments," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2, april 2007, pp. II–413–II–416.
- [10] D. R. Hardoon, S. R. Szedmak, and J. R. Shawe-taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, dec 2004.
- [11] J. Su, Z. Zhu, A. Srivastava, and F. Huffer, "Detection of Shapes in 2D Point Clouds Generated from Images," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ser. ICPR '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 2640–2643.
- [12] T. Garlipp and C. H. Müller, "Detection of linear and circular shapes in image analysis," *Comput. Stat. Data Anal.*, vol. 51, no. 3, pp. 1479–1490, Dec. 2006.
- [13] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [14] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [15] N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62 –66, jan. 1979.
- [16] J. B. T. M. Roerdink and A. Meijster, "The watershed transform: definitions, algorithms and parallelization strategies," *Fundam. Inf.*, vol. 41, no. 1-2, pp. 187–228, Jan. 2000.
- [17] F. Meyer, "Color image segmentation," in *Image Processing and its Applications, 1992., International Conference on*, 1992, pp. 303 – 306.
- [18] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," vol. 2. Los Alamitos, CA, USA: IEEE Computer Society, Aug. 1999, pp. 246–252.
- [19] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, jan 1972.
- [20] J. Illingworth and J. Kittler, "A survey of the hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87 – 116, 1988.
- [21] H. Kälviäinen, P. Hirvonen, L. Xu, and E. Oja, "Comparisons of probabilistic and non-probabilistic hough transforms," in *Proc. of 3rd European Conference on Computer Vision ECCV'94*, 1994, pp. 351–360.
- [22] A. Schmeder, A. Freed, and D. Wessel, "Best practices for open sound control," in *Linux Audio Conference*, Utrecht, NL, 01/05/2010 2010.
- [23] Q. Zhao, G. Zhang, R. Wood, and Z. Luo, "Video Based Real-Time Pedestrian Detection on Zebra Cross," in *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*, oct. 2009, pp. 1–4.
- [24] J. Jacques Junior, S. Musse, and C. Jung, "Crowd analysis using computer vision techniques," *Signal Processing Magazine, IEEE*, vol. 27, no. 5, pp. 66–77, sept 2010.